

DNN在CTR预估上的应用

1.CTR预估

CTR预估是计算广告中最核心的算法之一，那么CTR预估是指什么呢？简单来说，CTR预估是对每次广告的点击情况做出预测，预测用户是点击还是不点击。具体定义可以参考 CTR。CTR预估和很多因素相关，比如历史点击率、广告位置、时间、用户等。CTR预估模型就是综合考虑各种因素、特征，在大量历史数据上训练得到的模型。CTR预估的训练样本一般从历史log、离线特征库获得。样本标签相对容易，用户点击标记为1，没有点击标记为0。特征则会考虑很多，例如用户的人口学特征、广告自身特征、广告展示特征等。这些特征中会用到很多类别特征，例如用户所属职业、广告展示的IP地址等。一般对于类别特征会采用One-Hot编码，例如职业有三种：学生、白领、工人，那么会会用一个长度为3的向量分别表示他们： $[1, 0, 0]$ 、 $[0, 1, 0]$ 、 $[0, 0, 1]$ 。可以这样会使得特征维度扩展很大，同时特征会非常稀疏。目前很多公司的广告特征库都是上亿级别的。

2.DNN

深度神经网络(DNN)近年来在图像、语音、自然语言等领域大放异彩，特别是在图像分类、语音识别、机器翻译方面DNN已经超过人，精度已经达到商业应用程度。不过，DNN在CTR预估这种场景的应用却仍在摸索中。图像、语言、自然语言领域的数据一般是连续的，局部之间存在某些结构。比如，图像的局部与其周围存在着紧密的联系；语音和文字的前后存在强相关性。但是CTR预估的数据如前面介绍，是非常离散的，特征前后之间的关系很多是我们排列的结果，并非本身是相互联系的。

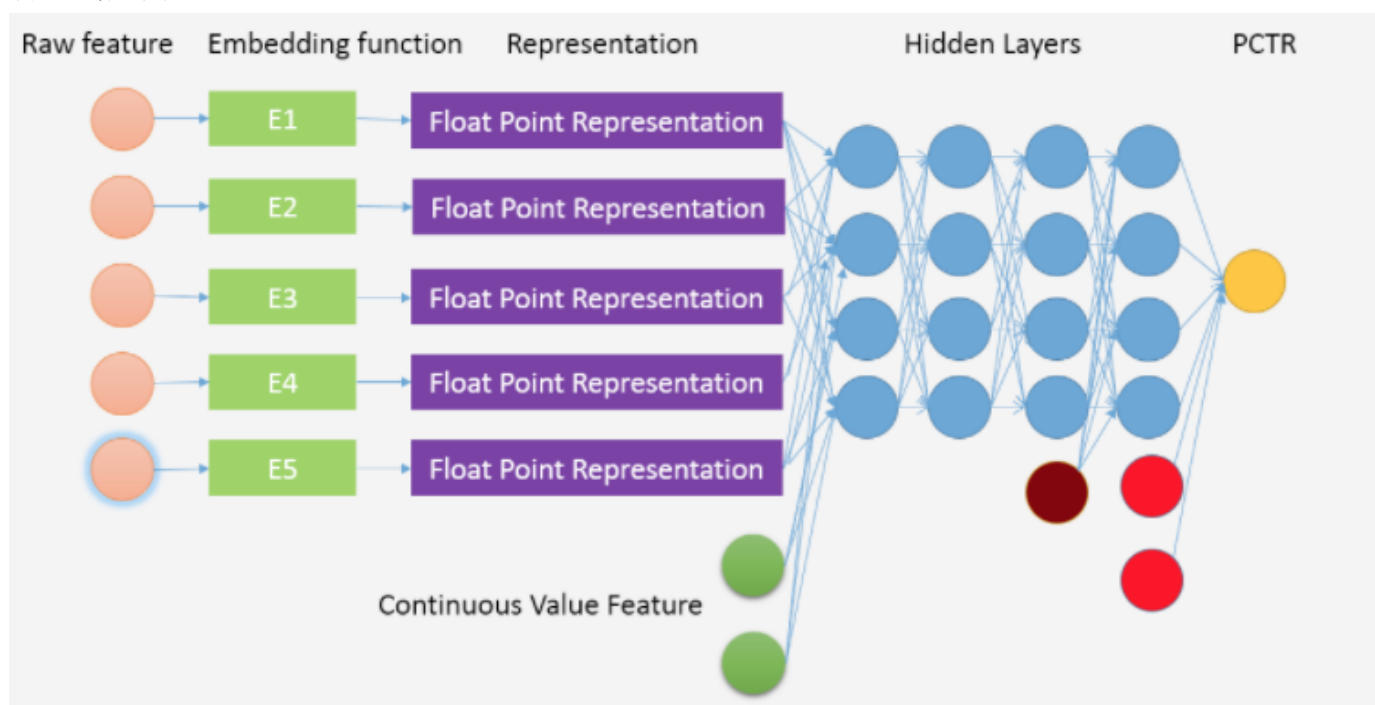
3.Embedding

Neural Network是典型的连续值模型，而CTR预估的输入更多时候是离散特征，因此一个自然的想法就是如何将离散特征转换为连续特征。如果你对词向量模型熟悉的话，

可以发现之间的共通点。在自然语言处理(NLP)中，为了将自然语言交给机器学习中的算法来处理，通常需要首先将语言数学化，词向量就是用来将语言中的词进行数一种方式。

一种最简单的词向量方式是one-hot，但这么做不能很好的刻画词之间的关系(例如相似性)，另外数据规模会非常大，带来维度灾难。因此Embedding的方法被提出，基本思路是将词都映射成一个固定长度的向量(向量大小远小于one-hot编码向量大些)，向量中元素不再是只有一位是1，而是每一位都有值。将所有词向量放在一起就是一个词向量空间，这样就可以表达词之间的关系，同时达到降维的效果。

既然Embedding可以将离散的词表达成连续值的词向量，那么对于CTR中的类别特征也可以使用Embedding得到连续值向量，再和其他连续值特征构成NN的输入。下图就是这种思路的表达。



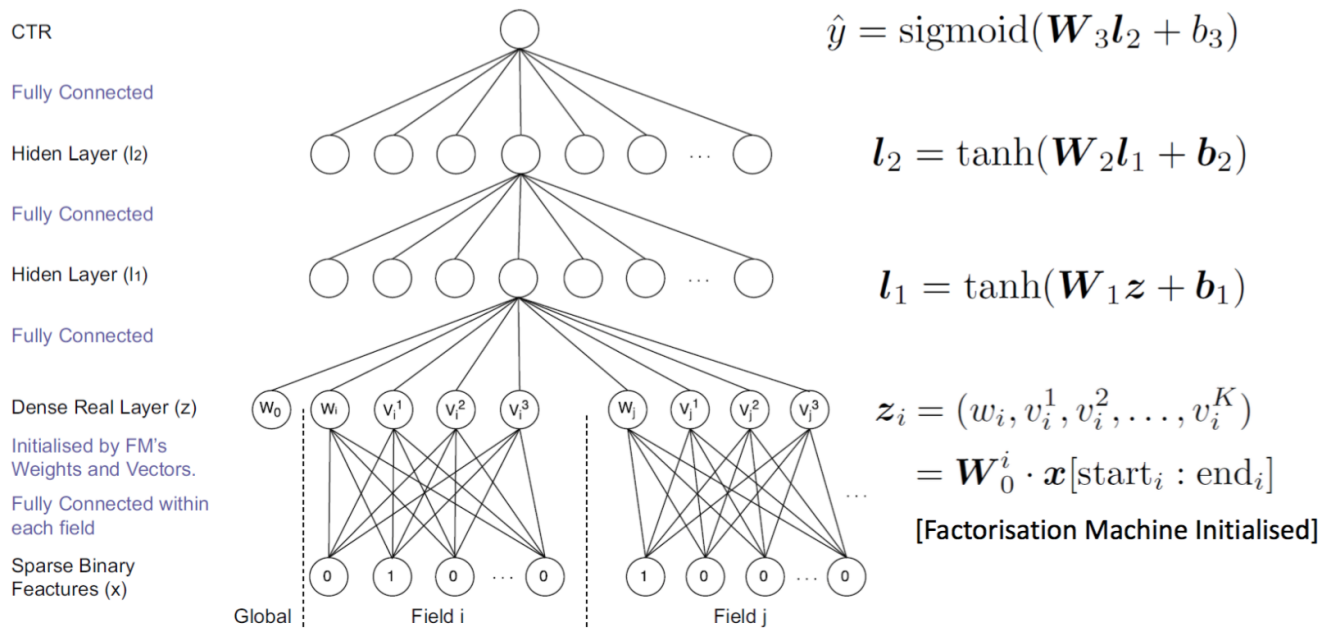
因此问题的关键就是采用何种Embedding技术将离线特征转换到离线空间。

3.1 FM Embedding

Factorization Machine是近年来在推荐、CTR预估中常用的一种算法，该算法在LR的基础上考虑交叉项，如下面公式所示：

$$y_{\text{FM}}(\mathbf{x}) := \text{sigmoid}\left(\underbrace{w_0 + \sum_{i=1}^N w_i x_i}_{\text{Logistic Regression}} + \underbrace{\sum_{i=1}^N \sum_{j=i+1}^N \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j}_{\text{Feature Interactions}}\right)$$

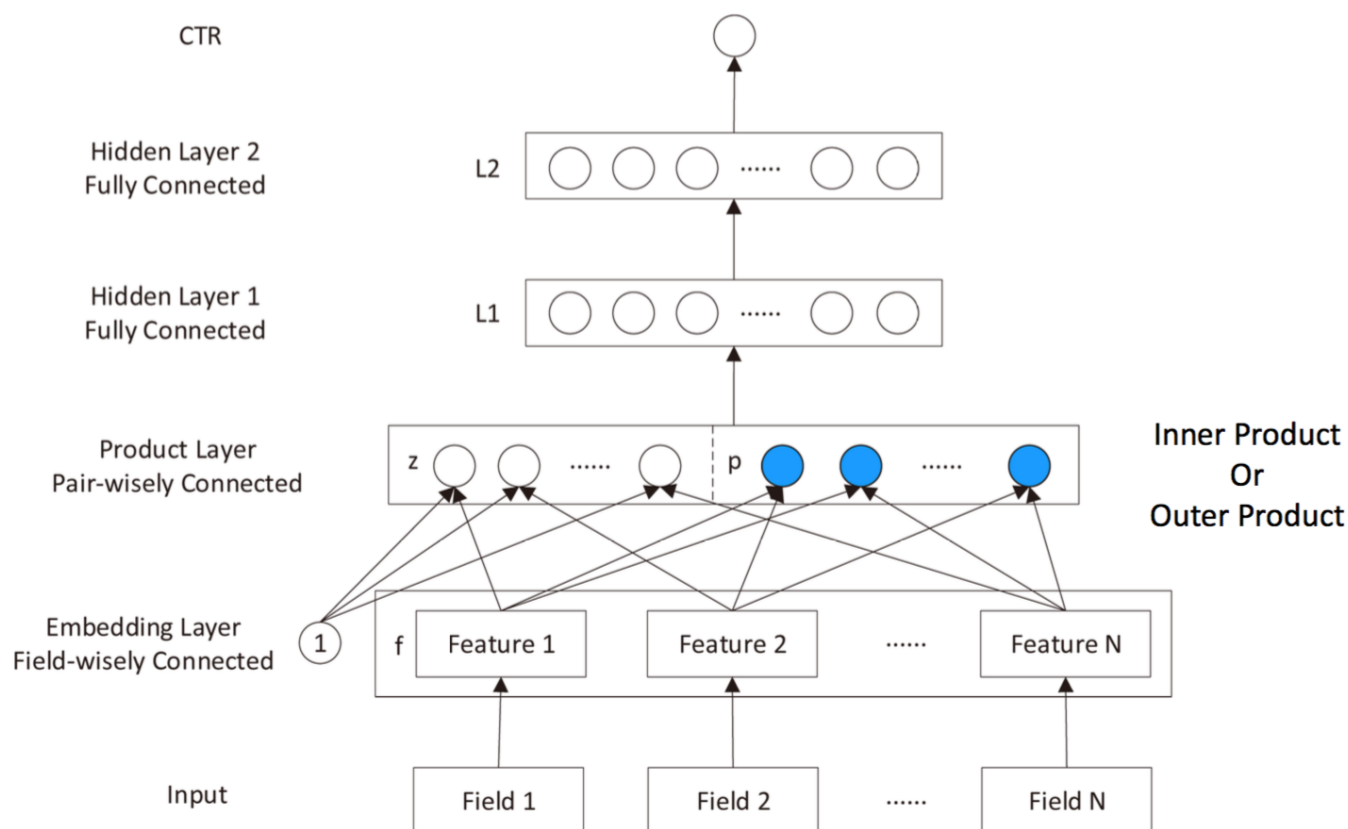
FM在后半部分的交叉项中为每个特征都分配一个特征向量 \mathbf{V} ，这其实可以看作是一种Embedding的方法。Dr.Zhang在文献[1]中提出一种利用FM得到特征的embedding向量并将其组合成dense real层作为DNN的输入的模型，FNN。FNN模型的具体设计如下：



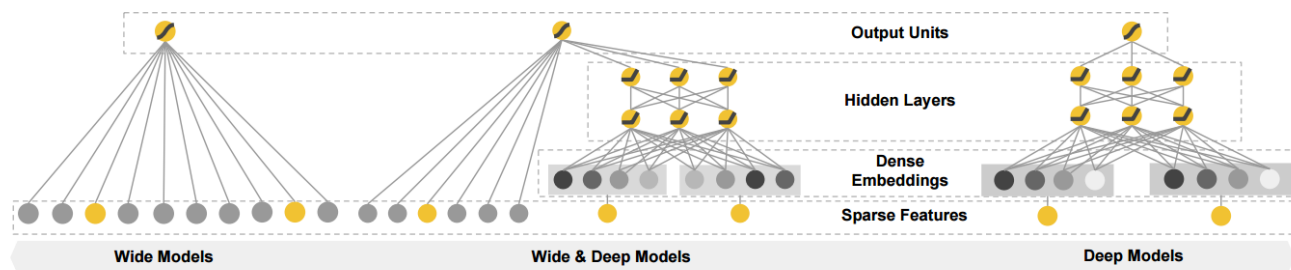
Dr.Zhang在模型中做了一个假设，就是每个category field只有一个值为1，也就是每个field是个one-hot表达向量。field是指特征的种类，例如将特征occupation one-hot之后是三维向量，但这个向量都属于一个field，就是occupation。这样虽然离散化后的特征有几亿，但是category field一般是几十到几百。模型得到每个特征的Embedding向量后，将特征归纳到其属于field，得到向量 \mathbf{z} ， \mathbf{z} 的大小就是 $1 + \# \text{fields} * \# \text{embedding}$ 。 \mathbf{z} 是一个固定长度的向量之后再在上面加入多个隐藏层最终得到FNN模型。

Dr.Zhang在FNN模型的基础上又提出了下面的新模型PNN。PNN和FNN的主要不同在于除了得到 \mathbf{z} 向量，还增加了一个 \mathbf{p} 向量，即Product向量。Product向量由每个category field的feature vector做inner product 或则 outer product 得到，作者认为这样做有助于

特征交叉。另外PNN中Embedding层不再由FM生成，可以在整个网络中训练得到。



3.2 NN Embedding



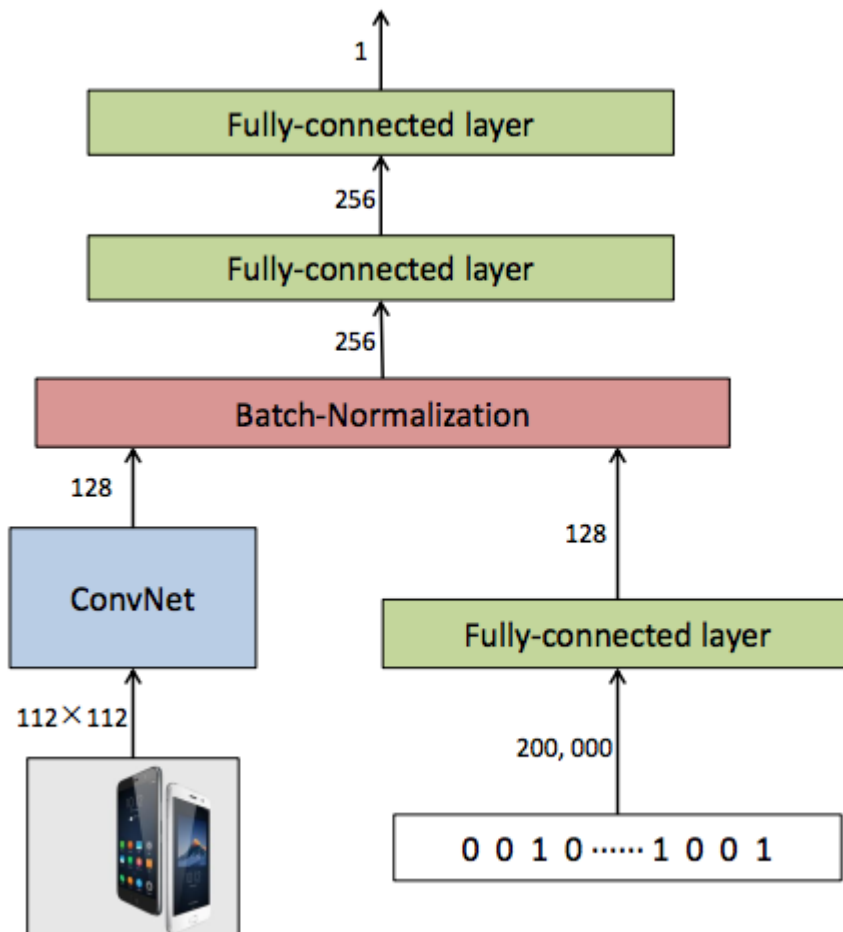
Google团队最近提出Wide and Deep Model。在他们的模型中，Wide Models其实就是LR模型，输入原始的特征和一些交叉组合特征；Deep Models通过Embedding层将稀疏的特征转换为稠密的特征，再使用DNN。最后将两个模型Join得到整个大模型，他们认为模型具有memorization and generalization特性。Wide and Deep Model中原始特征既可以是category，也可以是continue，这样更符合一般的场景。另外Embedding层是将每个category特征分别映射到embedding size的向量，如他们在TensorFlow代码中所示：

```
deep_columns = [  
    tf.contrib.layers.embedding_column(workclass, dimension=8),
```

```
tf.contrib.layers.embedding_column(education, dimension=8),
tf.contrib.layers.embedding_column(gender, dimension=8),
tf.contrib.layers.embedding_column(relationship, dimension=8),
tf.contrib.layers.embedding_column(native_country, dimension=8),
tf.contrib.layers.embedding_column(occupation, dimension=8),
age, education_num, capital_gain, capital_loss, hours_per_week]
```

4.结合图像

目前很多在线广告都是图片形式的，文献[4]提出将图像也做为特征的输入。这样原始特征就分为两类，图像部分使用CNN，非图像部分使用NN处理。其实这篇文章并没有太多新颖的方法，只能说多了一种特征。对于非图像特征，作者直接使用全连接神经网络，并没有使用Embedding。



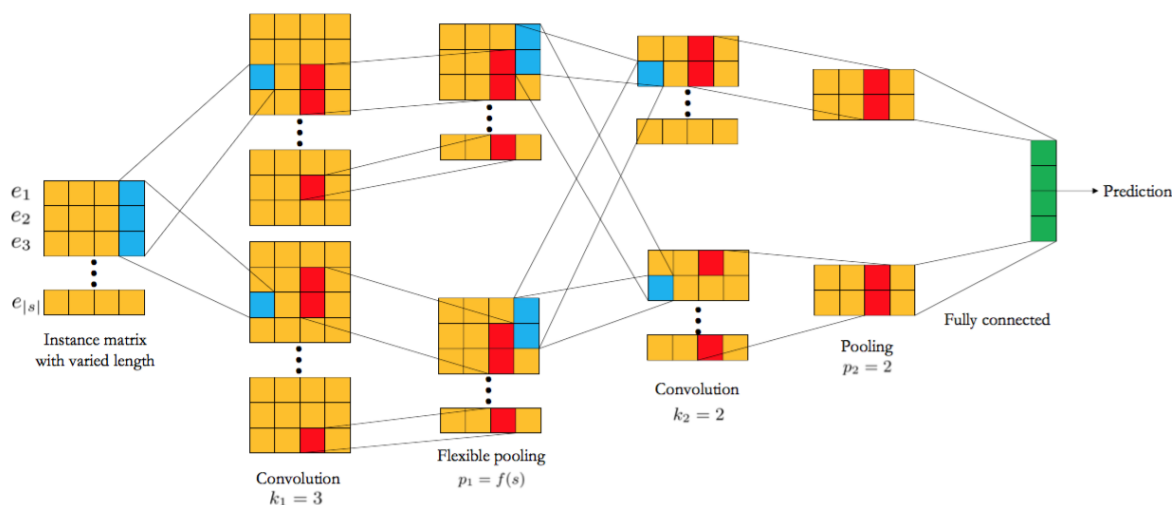
5.CNN

CNN用于提取局部特征，在图像、NLP都取得不错的效果，如果在CTR预估中使用却是个难题。我认为最大的困难是如何构建对一个样本构建如图像那样的矩阵，能够具有局

部联系和结构。如果不能构造这样的矩阵，使用CNN是没有什么意思的。文献[5]是发表在CIKM2015的一篇短文，文章提出对使用CNN来进行CTR预估进行了尝试。广告展示(single ad impression)包括：element = (user; query; ad, impression time, category, device type, etc) 用户是否点击一个广告与用户的历史ad impression有关。这样，一个样本将会是(s, label)，s由多条l组成(数目不定)

$$\mathbf{s} = \begin{bmatrix} \vdots & \vdots & \vdots \\ \mathbf{e}_1 & \cdots & \mathbf{e}_n \\ \vdots & \vdots & \vdots \end{bmatrix}.$$

作者提出CCPM模型处理这样的数据。每个样本有n个element，对每个element使用embedding 得到定长为d的向量 $e_i \in R^d$ ，再构造成一个矩阵 $\mathbf{s} \in R^{d \times n}$ ，得到s矩阵之后就可以套用CNN，后面的其实没有太多创新点。



6.RNN

考虑搜索场景下的CTR预估，如果考虑历史信息，如可以将一个用户的历史ad impression构成一个时间序列。RNN非常适合时间序列的场景，如语言建模等。这篇发表在AAAI2014将RNN模型引入CTR预估。作者首先在数据集上验证了用户的点击行为与之前的ad impression历史有关联：

- 如果用户在之前的impression很快离开广告页面，那么将会在接下来一段时间内不会点击类似的广告
- 如果用户最近有过与广告相关的查询，那么接下来点击相关广告的可能性会大幅提升
- 前面的两种行为还可能随着间隔时间的增加而不是那么相关

当前关联不止这些，而且人工难以刻画，需要模型来自动提取。RNN模型对此类问题非常适用，作者的主要工作是将数据集构造造成适合RNN的输入（即对用户的历史ad impression根据时间排序），对模型本身并没有改进。

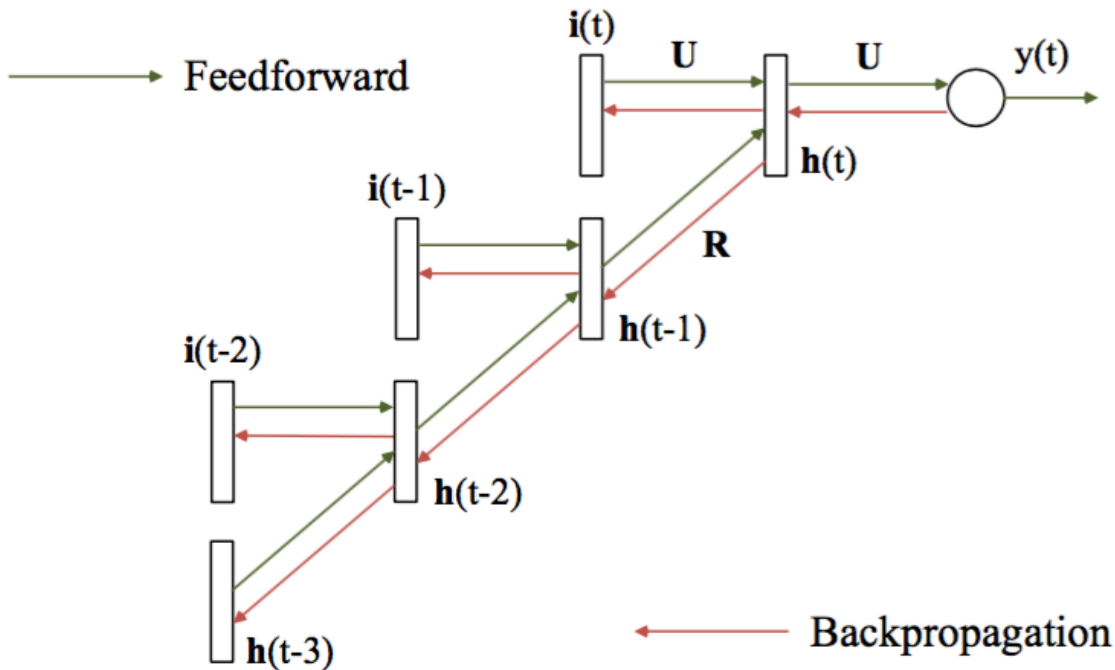


Figure 4: RNN training process with BPTT algorithm. Unfolding step is set to 3 in this figure.

参考文献

1. Deep Learning over Multi-field Categorical Data – A Case Study on User Response Prediction
2. Product-based Neural Networks for User Response Prediction
3. Wide & Deep Learning for Recommender Systems
4. Deep CTR Prediction in Display Advertising
5. A Convolutional Click Prediction Model
6. <http://www.52cs.org/?p=1046>


7. <http://techshow.ctrip.com/archives/1149.html>

8. <http://tech.meituan.com/deep-understanding-of-ffm-principles-and-practices.htm>

9. Sequential Click Prediction for Sponsored Search with Recurrent Neural Networks



Written by 沈成光

 Share on Twitter

 Share on Facebook

 Share on Google+

Updated March 16, 2017

[about.me](#)